

早稻氮磷钾施肥类别归属的贝叶斯判别方法研究

章明清¹, 李娟¹, 许文江², 孔庆波¹, 姚宝全³

(1 福建省农业科学院土壤肥料研究所, 福建福州 350013; 2 福建省亚热带植物研究所, 福建厦门 361006;
3 福建省农田建设与土壤肥料技术推广总站, 福建福州 350003)

摘要:【目的】在测土配方施肥中,“测土”是实现合理施肥微观指导的关键技术手段。本研究针对高度分散经营的耕地代表性土壤采集难、测试费用高、花费时间长等“测土”难题,探讨早稻氮磷钾施肥类别归属的统计模式识别技术。【方法】利用闽东南 80 个早稻氮磷钾田间肥效试验资料作为样本,采用欧氏距离-离差平方和法进行系统聚类分析,建立区域早稻施肥类别。在此基础上,根据田块氮磷钾施肥量及其产量结合地力信息等数量指标,应用贝叶斯判别准则探讨具体田块的施肥类别归属。【结果】在确保空白区产量和平衡施肥产量均值在两两施肥类别间具有统计显著差异的前提下,闽东南早稻最多可划分为 6 个氮磷钾施肥类别。多元统计检验表明,6 个施肥类别的类内协方差矩阵差异不全相等,据此建立了各个施肥类别的贝叶斯判别函数。结果表明,原始试验数据标准化处理可大幅度提高判别准确性,监测样本的回代误判率和交叉误判率分别只有 1.2% 和 1.3%,预留 6 个试验点的 84 个处理的氮、磷、钾施肥量和产量试验资料的施肥类别归属的判别正确率达到 81.0%。【结论】统计模式识别原理及其贝叶斯判别准则可用于早稻氮磷钾施肥类别归属决策,判别准确性能够满足推荐施肥的精度要求。

关键词: 早稻; 氮磷钾; 施肥类别; 贝叶斯判别; 类别归属

Bayesian discriminating analysis on category attribution of nitrogen, phosphorus and potassium fertilization for early rice

ZHANG Ming-qing¹, LI Juan¹, XU Wen-jiang², KONG Qing-bo¹, YAO Bao-quan³

(1 Soil and Fertilizer Institute, Fujian Academy of Agricultural Sciences, Fuzhou 350013, China; 2 Fujian Institute of Subtropical Plants, Xiamen 36000, China; 3 Fujian Cropland Construction and Soil and Fertilizer Station, Fuzhou 350003, China)

Abstract: 【Objectives】In soil testing and formulated fertilization, soil testing is the key to realize rational fertilization through practical guidance. However, collection of representative soil samples is often difficult because of highly decentralized farmland management, and soil sample analysis is cost and time-consuming. Therefore, statistical pattern recognition techniques were studied in this paper to explore category attribution of nitrogen (N), phosphorus (P) and potassium (K) fertilization without soil testing for early rice. 【Methods】Data from eighty field experiments in southeast of Fujian Province, China, were used in this study. Based on the response of early rice to N, P and K fertilizers, the paddy fields were divided into regional fertilization categories, using clustering analysis method of Euclidean distance-sum of squares of deviations. Then the NPK fertilization category for a field was calculated based on the statistical pattern recognition principle, the application rates of N, P and K fertilizers and the outputs. 【Results】On condition of ensuring that the average yield of the blank area and that of balanced fertilization had statistical significant differences between any two fertilization categories, the 80 paddy fields were divided clearly into six fertilization categories. Multivariate statistics showed that the differences of covariance matrix of six categories were not all equal, and the Bayesian discrimination function of each category was established based on that. The standardization of the original data greatly improved the discrimination accuracy, with the back substitution misjudgment rate and cross misjudgment rate of training

收稿日期: 2016-08-29 接受日期: 2017-03-20

基金项目: 国家自然科学基金项目(31572203); 福建省农业科学院 PI 项目(2016PI-31); 农业部测土配方施肥(2011~2015)资助。

作者简介: 章明清(1963—), 福建永春人, 博士, 研究员, 主要从事作物施肥原理与技术研究。E-mail: Zhangmq2001@163.com

samples of only 1.2% and 1.3% respectively, and the category discrimination accuracy of the 84 treatments in the 6 reserved experimental sites reached 81.0%. **【 Conclusions 】** Statistical pattern recognition principle and its Bayesian discriminate analysis method may provide an effective technical approach for the attribution decisions of N, P and K fertilization category of early rice without soil testing, and its results could meet the accuracy requirements of fertilizer recommendation.

Key words: early rice; nitrogen; phosphorus; potassium; fertilization category; Bayesian discrimination; category attribution

近 10 年来,我国在测土配方施肥技术推广应用过程中,利用“3414”设计开展了数以万计的作物氮磷钾田间肥效试验,数量之多和范围之广是罕见的。如何正确利用这些田间肥效试验结果来指导作物合理施肥是一个大问题。根据“3414”试验设计特点,一个试验点结果可分别建立三个一元肥效模型、三个二元肥效模型和一个三元肥效模型。肥效模型的优点是直观和计量准确,但其微观指导功能较弱。为了弥补肥效模型在应用上的不足,一些学者将土壤有效养分含量等地点变量引入到模型中^[1,2],或者建立土测值与肥效模型推荐施肥量的回归关系式^[3-5],从而实现了微观指导功能。显然,这种改进是建立在田间试验基础上,通过田间试验获取模型参数,在具体应用时都需要测土这一技术环节。但是,在广大农村的取土和测土过程中,常遇到代表性土样采集难、测试费用高和耽误农时等难题,制约了计量施肥技术的推广普及^[6]。因此,在田间肥效试验结合测土的研究基础上,探讨高效和准确的推荐施肥技术具有重要的实用价值,可对测土配方施肥起到有益补充。

推荐施肥技术如何实现既高效快速又有较高准确度呢?统计模式识别原理和技术为我们提供了有益的启发。该技术已在生物技术^[7-8]、作物品种鉴别^[9]、土壤分类^[10-13]、水土保持区划^[14]等领域得到应用,但在作物施肥类别归属判别上的应用还鲜见报道。为此,本文利用闽东南地区近年来完成的 86 个早稻“3414”设计氮磷钾田间肥效试验资料,在建立早稻氮磷钾施肥类别基础上,探讨区域内具体田块氮磷钾施肥类别归属的贝叶斯判别方法及其可行性,以期为最大限度地发挥众多“3414”田间试验资料在指导合理施肥中的作用提供一种新途径。

1 材料与方法

1.1 早稻氮磷钾肥效试验资料的收集整理

在近年来的测土配方施肥工作中,福建在福州

市、莆田市、泉州市和漳州市等闽东南沿海地区早稻上先后完成了 86 个“3414”设计的氮磷钾田间肥效试验。每个试验设 14 个处理,即: 1) $N_0P_0K_0$; 2) $N_0P_2K_2$; 3) $N_1P_2K_2$; 4) $N_2P_0K_2$; 5) $N_2P_1K_2$; 6) $N_2P_2K_2$; 7) $N_2P_3K_2$; 8) $N_2P_2K_0$; 9) $N_2P_2K_1$; 10) $N_2P_2K_3$; 11) $N_3P_2K_2$; 12) $N_1P_1K_2$; 13) $N_1P_2K_1$; 14) $N_2P_1K_1$ 。其中,“2”水平为试验前当地推荐施肥量,“0”水平表示不施肥,“1”水平和“3”水平的施肥量分别为“2”水平的 50% 和 150%,具体试验施肥量设计和实施方案与李娟等^[4]相同。

在 86 个试验资料中,取 80 个试验资料作为建立氮磷钾施肥类别的训练样本,按照不同地区预留 6 个试验点资料(表 1)用于检验判别结果可靠性。这 6 个试验资料分别来自闽侯县灰泥田、闽清县黄泥田、仙游县灰泥田和灰沙田、龙海市灰泥田以及平和县灰沙田的试验结果。基础土样采用常规方法^[15]测定土壤主要理化性状,其中,土壤 pH 为 4.7 ± 0.6 ,土壤有机质为 22.7 ± 1.0 g/kg,土壤碱解氮、Olsen-P 和速效钾含量分别为 120.7 ± 43.8 mg/kg、 33.9 ± 22.8 mg/kg 和 69.3 ± 32.7 mg/kg。每个试验点都有 14 个处理的氮、磷、钾施肥量和产量数据,即有 84 个待判样本,其中,处理 1) 无肥区产量水平变化幅度为 3315~7470 kg/hm²,处理 6) 平衡施肥产量变化幅度为 5355~10500 kg/hm²,基本涵盖了该区域早稻产量水平的可能范围。

1.2 氮磷钾施肥类别归属的判别原理

根据统计模式识别原理^[16],要进行未知施肥类别归属的判别分析,首先要在多点田间肥效试验基础上,构建研究区域的早稻施肥类别。这就要对该区域内的早稻肥效试验资料进行合理的定量分类。研究表明,对研究区域较小的作物氮磷钾肥效试验资料,欧氏距离-离差平方和法是最佳的系统聚类分析方法^[17],因而采用该法建立闽东南早稻氮磷钾施肥类别及其类特征肥效方程,并据此提出各施肥类别的推荐施肥量。

表 1 “3414” 预留试验点的施肥量及其产量试验数据

Table 1 Application rates of N, P and K fertilizers and the yields in the 14 treatments of six reserved experiments

试验点 Trial site	施肥量 Application rate			处理 Treatment													
	N	P ₂ O ₅	K ₂ O	N ₀ P ₀ K ₀	N ₀ P ₂ K ₂	N ₁ P ₂ K ₂	N ₂ P ₀ K ₂	N ₂ P ₁ K ₂	N ₂ P ₂ K ₂	N ₂ P ₃ K ₂	N ₂ P ₂ K ₀	N ₂ P ₂ K ₁	N ₂ P ₂ K ₃	N ₃ P ₂ K ₂	N ₁ P ₁ K ₂	N ₁ P ₂ K ₁	N ₂ P ₁ K ₁
T1	165	69	101	4295	4790	5815	5815	6290	6410	6370	3870	6070	5985	6310	6090	5730	5940
T2	150	60	90	3675	4050	5025	3975	5025	5415	5025	4950	5190	5280	5280	5175	5265	4950
T3	150	60	105	5345	5490	6513	6584	6792	6858	6663	6058	6697	6358	6107	6428	6446	6761
T4	165	55	110	5250	6065	7260	7200	7595	7745	7705	7160	7535	7465	7310	7125	7275	7340
T5	150	60	105	6150	6105	9750	6000	9510	10500	9510	6255	9510	9000	8505	8250	8505	8760
T6	150	60	105	7470	6225	8600	8300	8825	8876	8351	8576	9026	8576	8450	8675	8475	8745

注 (Note): 氮磷钾施肥量为“2”水平的施用量 The application rates of N, P and K are the fertilizer amount of “2” level in the treatments.

早稻施肥类别归属判别是指在构建区域早稻施肥类别后, 确定该区域内某个田块或几个相邻田块同种作物的施肥类别应属于这些已知类别中的哪一类。贝叶斯判别准则是统计模式识别中应用最广泛的模式归属判别方法, 是根据待判样本属于各个类别的条件概率值大小, 将其判归条件概率最大的那个类别。对多于 2 个以上类别的贝叶斯判别, 当各施肥类别的协方差矩阵相等时, 基于误判损失相等的贝叶斯判别函数^[16, 18]为:

$$W_j(x) = \mu_j^T \Sigma^{-1} x - \frac{1}{2} \mu_j^T \Sigma^{-1} \mu_j + \ln p_j, \quad j = 1, 2, \dots, k \quad (1)$$

判别准则为: 若 $W_j(x) = \max_{1 \leq j \leq k} \{W_j(x)\}, x \in G_j$ 。或者, 基于后验概率的贝叶斯判别函数^[16, 18]为:

$$d_j^2(x) = (x - \mu_j)^T \Sigma^{-1} (x - \mu_j) - 2 \ln p_j, \quad j = 1, 2, \dots, k \quad (2)$$

其判别准则为: 若 $d_j^2(x) = \min_{1 \leq j \leq k} \{d_j^2(x)\}, x \in G_j$ 。式中, x 为待判样本, μ_j 表示第 j 个施肥类别的氮、磷、钾施肥量及其产量指标均值, Σ^{-1} 表示施肥类别中各施肥类别的协方差矩阵的逆矩阵, p_j 表示第 j 个施肥类别的先验概率值, $W_j(x)$ 和 $d_j^2(x)$ 表示待判样本 x 在第 j 个施肥类别的判别计算值, G 表示施肥类别, k 表示区域氮磷钾施肥类别数。若各施肥类别总体的各指标均值和协方差矩阵未知时, 可用训练样本的相应均值和协方差矩阵代替。

当各施肥类别的协方差矩阵不全相等时, 基于后验概率的贝叶斯判别函数^[16, 18]则为:

$$d_j^2(x) = (x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j) + \ln |\Sigma_j| - 2 \ln p_j \quad (3)$$

这时的判别准则为: 若 $d_j^2(x) = \min_{1 \leq j \leq k} \{d_j^2(x)\}$,

$x \in G_j$ 。式中, Σ_j^{-1} 表示第 j 个施肥类别的协方差矩阵的逆矩阵, 其余符号意义同上。

1.3 数据标准化和协方差矩阵差异显著性检验

在建立判别函数时要用到的均值和协方差矩阵, 其数值大小受到数据量级和量纲的影响。因此, 在建立各施肥类别的判别函数前, 需对施肥量和产量原始数据进行标准化处理。计算公式为:

$$x^* = \frac{(x - \bar{x})}{s} \quad (4)$$

其中: x^* 表示 x 指标经标准化处理后的数值, \bar{x} 和 s 分别表示氮、磷、钾施肥量和产量各指标的平均值和标准差。

由于施肥类别间协方差矩阵相等或不全相等时的判别函数不同, 因而需要对各施肥类别的协方差矩阵进行差异显著性检验。假设有氮、磷、钾施肥量和产量 4 个指标观测值的 k 个施肥类别 $G_i (i = 1, 2, 3, \dots, k)$, 分别从中抽取样本容量为 n_i 的 k 个样本, 协方差矩阵分别为 S_1, S_2, \dots, S_k 。原假设为各施肥类别的协方差矩阵相等, 则统计量^[18-19]:

$$\xi = (1 - d)M \sim \chi^2(df) \quad (5)$$

其中,

$$d = \begin{cases} \frac{2p^2 + 3p - 1}{6(p + 1)(k - 1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{n - k} \right), & n_i \text{ 不全相等} \\ \frac{(2p^2 + 3p - 1)(k + 1)}{6(p + 1)(n - k)}, & n_i \text{ 全相等} \end{cases}$$

其中: $M = (n - k) \ln |S| - \sum_{i=1}^k (n_i - 1) \ln |S_i|,$

$S = \sum_{i=1}^k (n_i - 1) S_i / (n - k),$ 自由度 $df = p(p + 1)(k - 1) / 2,$

样本总数 $n = n_1 + n_2 + \dots + n_k$ 。对给定的概率值 a ，计算概率 $p = P(\xi > \chi^2 a^2(df))$ ，若 $p < a$ ，则拒绝原假设，否则接受原假设。在本文中， $p=4$ ， $k=6$ ， $n=80$ ， n_i 为各施肥类别的试验点数量。

1.4 系统聚类分析和施肥类别归属判别的简洁方法

构建施肥类别的系统聚类分析和施肥类别归属的贝叶斯判别分析均涉及到深奥数学原理和复杂计算，不可能通过人工演算或计算器计算来完成。本文聚类分析的具体计算过程调用 MATLAB 软件的 `pdist`、`linkage` 和 `dendrogram` 等功能函数完成。贝叶斯判别分析采用 MATLAB 软件提供的功能强大的 `classify` 函数来进行相关计算，具体使用方法可参阅相关专著^[19]。

考虑到现有参考文献大都未能准确地介绍 `classify` 函数的使用方法，根据本研究对此的学习和体会简要介绍如下，以供相关应用时参考。`classify` 函数的调用格式为：`[class, err, Poster] = classify(sample, training, group, type, prior)`。其中，`class` 给出待判样本归属的判别结果，`err` 是训练样本回代误判率，`Poster` 是待判样本归属的后验概率。`sample` 是指待判样本，`training` 是指已知类别归属的训练样本，`group` 是对应训练样本的类别编号，`type` 是指具体判别分析方法的选择项，包括 `linear`、`quadratic` 和 `mahal` 等，`prior` 是指先验概率的可选项，用来指定各类别的先验概率值。应特别注意，`type` 选项若选择 `linear` 或 `quadratic`，`classify` 函数将采用贝叶斯判别，若选用 `mahal` (马氏距离) 时则采用距离判别，此时

`Poster` 选项的输出项为空集，先验概率只能用来计算误判概率。上述相关输入和输出数据均以矩阵或向量的形式表达。MATLAB 软件没有提供交叉误判率的计算程序，要使用必须自己编写计算程序。

对于试验数据标准化处理，本文采用 MATLAB 软件的 `zscore` 函数，其调用格式为 `[S, mu, sig] = zscore(X)`。其中，`X` 表示原始试验数据矩阵，`S` 表示数据标准化处理结果，`mu` 和 `sig` 分别表示原始数据各观测指标的平均值和标准差。

2 结果与分析

2.1 早稻氮磷钾施肥类别的构建

为建立闽东南地区早稻氮磷钾施肥类别，参照毛达如等^[20]建立类特征肥效方程的思路，利用 80 个早稻“3414”设计的氮磷钾田间肥效试验资料，采用欧氏距离-离差平方和法对每个试验点的 14 个处理产量进行系统聚类分析，并对各施肥类别的处理 1) 空白区产量和处理 6) 平衡施肥产量进行均值差异显著性检验^[18]。

表 2 的结果显示，当 80 个试验资料分成 9 类或 10 类时，总有两类或两类以上的均值差异不显著，导致定量分类结果无效；分成 7 类或 7 类以上时，总有一类的试验点数只有一个，导致该类别因试验点过少而缺乏代表性。因此，有效分类数最多为 6 个施肥类别，此时每一个施肥类别至少有 7 个代表性试验点，同时确保了类与类之间的空白区产量均值和平衡施肥产量均值的差异达到统计显著水平。

表 2 早稻氮磷钾试验处理类别及类别间差异显著性分析 ($P < 0.05$)

Table 2 Category of the fertilization treatments and the paired significance analysis of the categories

分类数 Category	处理类别序号及其所含试验点数 (n) Fertilization category order and the trials number										两类间差异显著性统计 Statistics between any two categories	
	1	2	3	4	5	6	7	8	9	10	组合数 Total number	差异显著组合数 No. of significant difference
3	7	52	21								3	3
4	20	32	7	21							6	6
5	13	8	20	32	7						11	11
6	9	23	13	8	20	7					15	15
7	1	6	9	23	13	8	20				22	22
8	8	15	1	6	9	13	8	16			29	29
9	1	5	8	15	1	9	12	8	20		37	35
10	4	5	1	5	8	15	1	13	8	20	46	43

注 (Note): 表中数字为各试验点的处理代码 Data are the treatment codes in each trial.

结果表明, 依据空白区产量和平衡施肥产量水平, 6 个施肥类别可分别命名为“超高产田”、“高产田”、“中高产田”、“中产田”、“中低产田”和“低产田”等六种稻田施肥类型(表 3)。

根据各施肥类别对应试验点的试验结果, 分别建立氮磷钾三元二次多项式肥效模型(略)。统计表明, 6 个施肥类别的三元肥效模型的 F 值均达显著水平, 而且模型参数正负号满足水稻营养特性, 均为典型肥效模型^[21]。因此, 以每公斤 N 4.3 元、 P_2O_5 5 元、 K_2O 6 元和稻谷 2 元的当地市场均价为依据, 用边际产量导数法计算各施肥类别的推荐施肥量(表 3)。但在指导该区域早稻合理施肥时, 用户的具体田块的推荐施肥该归属于哪个施肥类别呢? 以往常用的方法是根据用户常年空白区产量或者预期施肥目标产量, 对照表 3 的结果进行推荐。但是, 在广大农村中空白区产量或施肥目标产量等指标难以

准确获取, 在应用上尚存在困难, 模式识别技术可使表 3 结果的实际应用更为简便和可靠。

2.2 早稻氮磷钾施肥类别的训练样本

根据统计模式识别原理^[6], 要对未知施肥类别的地块做出施肥决策, 首先应根据各施肥类别的相关试验资料, 提取各类别的特征样本作为训练样本。这样做的缘由是使后续建立决策准则更容易, 结果更稳定和易于理解。因此, 根据 80 个试验点在 6 个施肥类别中的归属, 针对同一施肥类别内各试验点的 14 个处理, 提取相同处理的氮、磷、钾施肥量及其对应产量均值, 作为该类别的训练样本, 结果见表 4。

根据“3414”试验设计方案, 表 4 的每个施肥类别有 14 个处理, 每个处理含有 4 个指标, 即: 氮、磷、钾施肥量及其产量。从表中可以看到, 施肥量数据在 10 的一次方和二次方量级之间, 而产量数据在 10 的三次方量级, 数值大小差异较大。为消

表 3 供试地块基础产量、目标产量及其推荐施肥量 (kg/hm^2)

Table 3 Paddy field yield levels, blank yields and the recommended fertilizer rates for the target yields in different categories of N, P and K fertilization recommendation for early rice

类别 Category No.	产量水平 Yield level	基础产量 Basic yield	目标产量 Target yield	最高施肥量 Max. fertilizer rate			最高产量 Max. yield	经济施肥量 Eco. fertilizer rate			经济产量 Eco. yield
				N	P_2O_5	K_2O		N	P_2O_5	K_2O	
				1	超高产 Super high	6809 ± 748		9411 ± 637	156	60	
2	高产 High	5573 ± 871	8074 ± 311	163	60	96	8054	140	51	73	7983
3	中高产 Mid & high	5461 ± 699	7406 ± 303	163	60	103	7414	130	48	80	7329
4	中产 Mid	4835 ± 436	6890 ± 259	175	67	113	6859	141	52	83	6759
5	中低产 Mid & Low	4403 ± 178	6327 ± 259	163	62	108	6330	125	50	85	6240
6	低产 Low	3538 ± 589	5661 ± 424	192	73	99	5664	152	58	77	5569

注 (Note): 基础产量为不施肥可获得的产量 Basic yield is the yield without fertilization.

表 4 各施肥类别的训练样本的氮、磷、钾施肥量和产量均值 (kg/hm^2)

Table 4 Average N, P and K fertilizer rates and early rice yields in the selected training samples from the six fertilization categories

类别 Category	施肥量 Fertilizer rate			处理 Treatment													
	N	P_2O_5	K_2O	$N_0P_0K_0$	$N_0P_2K_3$	$N_1P_2K_2$	$N_2P_0K_2$	$N_2P_1K_2$	$N_2P_2K_2$	$N_2P_3K_2$	$N_2P_2K_0$	$N_2P_2K_1$	$N_2P_2K_3$	$N_3P_2K_2$	$N_1P_1K_2$	$N_1P_2K_1$	$N_2P_1K_1$
1	156	66	104	6809	7219	8829	8021	9015	9411	9062	7938	9144	8927	8862	8753	8779	8782
2	158	56	107	5573	6033	7422	7118	7918	8074	7942	7264	7801	7613	7626	7315	7534	7701
3	160	61	104	5461	6012	6981	6742	7248	7406	7246	6117	7054	7073	7100	6938	6991	7097
4	159	64	106	4834	5225	6249	5931	6628	6890	6602	5681	6668	6624	6561	6221	6227	6470
5	161	64	104	4403	4840	5904	5087	6039	6327	5922	4372	5882	5968	5945	5811	5693	5772
6	150	60	97	3538	3701	4883	4222	5227	5661	5419	4527	5330	5169	5424	4922	4943	5153

注 (Note): $n = 6$. 表中数据未经标准化处理 Data are not standardized in the table. 氮磷钾施肥量为“2”水平的施用量 The application rates of N, P and K are the fertilizer amount of “2” level in the treatments.

除数据量级差异的影响,在建立判别函数前,需对 80 个试验资料的施肥量和产量原始数据根据 (4) 式进行标准化处理,然后利用处理后的数据分别建立 6 个施肥类别的判别函数。

2.3 早稻氮磷钾施肥类别归属的判别函数及其应用

将 80 个试验资料分成 6 个施肥类别时,根据 (5) 式的 ζ 统计量,各施肥类别的协方差矩阵差异显著性检验结果表明,自由度 df 等于 50,统计量 ζ 等于 376.97,大于临界值 $[\chi^2(df) = 67.51]$,表明各施肥类别间协方差矩阵不全相等。因此,需采用基于 (3) 式的贝叶斯判别函数。

为得到 (3) 式判别函数的具体表达式,如上所述,第一步,对 80 个试验资料分别计算各处理氮、磷、钾施肥量及其对应试验产量的平均值和标准差。第二步,利用第一步得到的各处理氮、磷、钾施肥量及其产量的平均值和标准差,分别计算表 4 中 6 个施肥类别的对应试验点 14 个处理中各处理对应的氮、磷、钾施肥量和产量的标准化数值。第三步,在此基础上,分别计算各施肥类别的氮、磷、钾施肥量和产量的均值、协方差矩阵及其逆矩阵。第四步,取各施肥类别的试验点数除以试验总数 (80 个),所得数值作为其先验概率值。最后,将所得数值分别代入 (3) 式判别函数式,分别得到 6 个施肥类别的判别函数表达式。

为便于理解,对表 4 的各施肥类别分别随机抽取一个处理作为样本,即氮、磷、钾施肥量及其对应产量,利用 MATLAB 软件的 classify 功能函数计算其后验概率值 (表 5)。结果表明,来自第 1、2、3、4、5、6 个施肥类别的样本分别归属于原所属施

肥类别的概率值为 0.9998、0.9999、0.9709、1.0000、0.9980、1.0000,即,判别结果正确地归属于它们原来的所属类别。

2.4 氮磷钾施肥类别归属准确性评价

模式识别准确性评估常用训练样本的回代误判率和交叉误判率指标。根据表 4 的 6 个施肥类别 80 个试验点的氮、磷、钾施肥量和产量数据,应用 MATLAB 软件提供的 classify 函数计算回代误判率,同时自编程序计算交叉误判率。结果表明,如果原始数据未经标准化处理就建立判别函数,回代误判率和交叉误判率则分别达到 34.5% 和 41.3%,判别准确性显然不能满足应用需要。但对原始数据进行标准化处理后建立判别函数,回代误判率和交叉误判率则分别下降到 1.2% 和 1.3%,完全满足推荐施肥的精度要求,说明训练样本数据经过标准化处理可大幅度提高判别准确性。

回代误判率和交叉误判率作为评价指标的缺点是可能高估了模式识别系统的性能。为此,进一步利用预留的 6 个试验点资料 (表 1) 作为待判样本来检验判别效果。首先参考当前多点试验资料汇总中土壤肥力等级划分的一般方法^[4-5],根据空白处理和平衡施肥处理的产量水平,与表 4 中各施肥类别的相应处理产量水平的接近程度,确定这 6 个试验点的施肥类别归属。然后,利用已建立的贝叶斯判别函数检验其类别归属是否一致。一般认为,判别准确性达到 80% 左右时^[18],该模式识别系统即可投入使用。表 6 结果表明,对包括氮、磷、钾施肥量和产量指标的 84 个处理的样本,平均判别准确率达到 81.0%。可见,贝叶斯判别的准确率达到了推荐施肥对精度的一般要求。

表 5 基于后验概率的贝叶斯判别函数的施肥类别归属的判别结果

Table 5 Recognition results of fertilization category attribution based on posterior probability of the Bayesian discriminant function

试验地点 Trial site	施肥量 (kg/hm ²) Nurtient rate			产量 (kg/hm ²) Yield	所属类别 Category	归属类别概率 Category attribution probability						判别归属 Attribution	
	N	P ₂ O ₅	K ₂ O			1	2	3	4	5	6		
													T1
T2	160.3	91.8	104.4	7246	2	0.0001	0.9999	0	0	0	0	0	2
T3	80.3	63.5	103.9	5904	3	0	0	0.9709	0	0.0291	0	0	3
T4	232.5	60.6	97.1	5424	4	0	0	0	1.0000	0	0	0	4
T5	79.3	64.2	53.2	6227	5	0	0	0.0020	0	0.9980	0	0	5
T6	156.4	65.6	52.0	9144	6	0	0	0	0	0	1.0000	0	6

注 (Note): 1、2、3、4、5、6 代表所属类别 1, 2, 3, 4, 5 和 6 are the categories the sites attributed. 氮磷钾施肥量为“2”水平的施用量 The application rates of N, P and K are the fertilizer amount of “2” level in the treatments.

表 6 预留 6 个试验点贝叶斯判别函数判别的施肥类别

Table 6 Fertilization categories of the reserved six trial sites attributed by the Bayesian discriminant function

试验点 Trial site	观测结果 Monitored result		判别结果 Discriminant result		
	样本数 Sample No.	所属类别 Fertilization category	归入所属类别样本数 Sample No. in the category	归入其它类别样本数 Sample No. in other category	平均准确率 (%) Average accuracy
T1	14	3	11	3	81.0
T2	14	4	13	1	
T3	14	5	14	0	
T4	14	1	11	3	
T5	14	6	10	4	
T6	14	6	9	5	

3 结论与讨论

3.1 测土施肥技术的适用性

土壤测定与推荐施肥技术经过 20 世纪四十年代至六十年代这一段时间的发展, 已奠定了良好的科学基础, 并确立了一整套国际公认与通用的工作方法^[22-23], 在农业规模化生产的欧美发达国家无不行之有效。我国在测土配方施肥技术发展过程中, 于 1990 年前后分别提出了“先测土后效应”和“先效应后测土”两大技术路线^[24-25], 要实现合理施肥的微观指导同样必须测土。毫无疑问, “测土”对于单个试验点的深入研究非常有效, 能够让人们对耕地土壤肥力状况、作物需肥规律和推荐施肥三者间关系有深入的认识和掌握。然而, 对面上的多点试验资料在推荐施肥中应用的深入探讨和规律总结则是其弱点, 这也是为什么众多“3414”试验资料至今未能被很好地归纳总结的一个关键技术原因。

同时, 在推广应用上, 这种源自欧美发达国家规模化农业生产条件下的测土推荐施肥技术, 面对我国耕地高度分散经营的农业生产实际时, 在取土、测土过程中经常会遇到目前还难以克服的难题^[6]。主要体现在两个方面: 1) 由于耕地高度分散经营和复种指数高, 农户间栽培管理水平存在差异, 造成土壤肥力空间变异大, 代表性土样采集难; 2) 为了获得较高精度的推荐施肥量, 必须加大土样采样密度, 结果造成测土费用高、耗费时间长和常误农时等问题。实际上, 这种状况就如同在火车站需要高效快速进行旅客身份识别, 通过对旅客采集生物样品进行 DNA 测序来鉴别身份的技术, 虽然结果准确和无异议, 但在这种场合并不实用, 而人脸模式识别技术则满足了准确和快速通关的客观需要。

因此, 在推荐施肥技术领域, 仅仅停留在通过

“测土”技术手段深入研究和掌握土壤肥力状况、作物需肥规律和推荐施肥三者间关系是不够的, 我们还需要在此基础上, 应用高度发达的信息技术和计算机技术对研究结果进行深度总结和挖掘, 提出普通用户就能使用的高效、快速和准确的肥效参数识别和推荐施肥新技术, 才能满足广大农户的实际应用需要。

3.2 统计模式识别技术的应用

统计模式识别理论和技术的发展和成熟, 为我们解决“测土”难题提供了一个新思路。在构建了区域作物施肥类别后, 对于该区域内某个田块或几个临近田块的推荐施肥应属于哪个施肥类别呢? 这就涉及到施肥类别归属的判别问题。根据统计模式识别理论^[16], 解决的办法是, 先对区域内各个施肥类别的相关观测指标进行统计分析, 确定每个施肥类别的统计特征, 再把未知施肥类别田块的相关指标与这些已知施肥类别的统计特征进行比较, 把它们归类到统计特征最相近的一个已知施肥类别中去, 进而提出施肥建议。因此, 可以采用具体地块上一年度或前 2~3 年的相同作物氮、磷、钾施肥量与产量, 结合地块肥力信息等数量指标, 作为施肥类别归属判别和推荐施肥的依据。

这一技术路线的准确性如何是人们最为关心的问题。本文初步研究表明, 这个问题应从应用层面和技术层面两个方面来考虑。在应用层面, 如果用户在施肥识别系统使用过程中逐渐养成记录田间资料档案的习惯, 由此得到的数据质量将明显提高。在技术层面, 在确保两类别间的关键指标均值具有显著差异的前提下, 将闽东南地区的早稻划分为 6 个施肥类别, 训练样本的回代误判率和交叉误判率平均只有 1% 左右, 预留的 84 个处理的氮、磷、钾

施肥量及其产量试验资料的施肥类别归属的判别正确率达到 81.0%。但是, 如果原始数据未经标准化处理就建立判别函数, 回代误判率和交叉误判率则分别高达 34.5% 和 41.3%。结果表明, 数据标准化处理是提高判别正确率的重要手段。

判别分析的另一种常用方法是距离判别法^[18], 利用该法和本文的 80 个试验资料作为训练样本来检验表 1 预留资料的判别准确性, 结果判别正确率只有 69.0%, 显示不同方法的判别准确性有较大差别。因此, 选择合适的判别方法是提高判别准确性的另外一个重要途径。由于统计模式识别的具体数学方法很多^[6], 在将来的工作中, 我们应根据不同区域的田间肥效试验资料, 深入探讨不同方法的专业适用性和可行性。此外, 提高判别准确率的另一个关键技术是正确建立作物施肥类别。

3.3 区域作物施肥类别的构建

统计模式识别技术在推荐施肥中应用的前提是正确地建立区域作物施肥类别。聚类分析是定量分类的最常用方法, 但聚类结果的谱系图只给出了各个试验点的亲疏关系, 本身并没有给出具体分类。本研究表明, 闽东南地区早稻氮磷钾施肥类别的有效分类数为 6 个。这种以施肥量和早稻产量为指标依据的施肥类别划分方法具有简单实用的优点。但是, 要了解不同施肥类别的土壤养分状况, 还需要结合土壤测定。

区域作物施肥类别应该划分多少类? 以往大都根据专业知识结合个人经验划分为 3~6 个类别, 然后按照不同类别所属试验点的田间试验资料进行归纳总结^[4-5]。显然, 这种方法带有明显的人为因素的影响。本研究表明, 施肥类别数取决于: 1) 类别的代表性, 即某个类别的试验数不能太少; 2) 类别间的关键指标均值差异达到统计显著水平; 3) 预留样本类别归属的判别正确率达到 80% 以上^[18]。这样, 我们就能将区域作物施肥类别划分建立在严格的定量分类基础上, 使每个施肥类别都具有鲜明的统计学意义, 这是提高施肥类别归属判别准确性的一个关键点。

本研究所涉及的闽东南地区, 在农业生产条件和早稻栽培管理水平等方面差异不大, 因而在建立早稻施肥类别时仅考虑氮磷钾施肥量及其产量作为指标依据, 就能达到 81.0% 的判别准确率。可以设想, 如果研究所涉及的地理区域较大, 导致生产条件和生产水平有明显差异时, 为使判别准确率

达到满意水平, 在确定施肥类别时, 土壤肥力信息、生态环境条件等地点变量就将变得很重要而不可忽略。

参 考 文 献:

- [1] 杨卓亚, 毛达如, 黄金龙, 卢志光. 基于年景变化的施肥决策[J]. 北京农业大学学报, 1995, 21(Suppl.): 19-22.
Yang Z Y, Mao D R, Huang J L, Lu Z G. Fertilization strategy for different crop yield year type[J]. Acta Agriculturae Universitatis Pekinensis, 1995, 21(Suppl.): 19-22.
- [2] 耿兴元, 毛达如, 曹一平. 土壤肥力模糊量化评价(判)系统的建立[J]. 北京农业大学学报, 1995, 21(Suppl.): 23-28.
Geng X Y, Mao D R, Cao Y P. A fuzzy system for quantitative evaluation of soil fertility[J]. Acta Agriculturae Universitatis Pekinensis, 1995, 21(Suppl.): 23-28.
- [3] 金耀青, 张中原. 配方施肥方法及其应用[M]. 沈阳: 辽宁科学技术出版社, 1993: 35-158.
Jin Y Q, Zhang Z Y. The method of formula fertilization and its application[M]. Shenyang: Liaoning Science and Technology Press, 1993: 35-158.
- [4] 李娟, 章明清, 孔庆波, 等. 福建早稻测土配方施肥指标体系研究[J]. 植物营养与肥料学报, 2010, 16(4): 938-946.
Li J, Zhang M Q, Kong Q B, et al. Soil testing and formula fertilization index for early rice in Fujian Province[J]. Plant Nutrition and Fertilizer Science, 2010, 16(4): 938-946.
- [5] 杨俐苹, 自由路, 王贺, 等. 测土配方施肥指标体系建立中“3414”试验方案应用探讨—以内蒙古海拉尔地区油菜“3414”试验为例[J]. 植物营养与肥料学报, 2011, 17(4): 1018-1023.
Yang L P, Bai Y L, Wan H, et al. Application of “3414” field trial design for establishing soil testing and fertilizer recommendation index[J]. Plant Nutrition and Fertilizer Science, 2011, 17(4): 1018-1023.
- [6] 侯彦林, 郭喆, 任军. 不测土条件下半定量施肥原理和模型评述[J]. 生态学杂志, 2002, 21(4): 31-35.
Hou Y L, Guo Z, Ren J. Summarization of principles and models for semi-quantitative fertilization without soil testing[J]. Chinese Journal of Ecology, 2002, 21(4): 31-35.
- [7] Kim J W, Ahn Y J, Lee K C, et al. A classification approach for genotyping viral sequences based on multidimensional scaling and linear discriminant analysis[J]. BMC Bioinformatics, 2010, 11: 434.
- [8] Zhao X J, Pei Z Y, Liu J, et al. Prediction of nucleosome DNA formation potential and nucleosome positioning using increment of diversity combined with quadratic discriminant analysis[J]. Chromosome Research, 2010, 18: 777-785.
- [9] Li X L, Yi S L, He S L. Identification of pummelo cultivars by using Vis/NIR spectra and pattern recognition methods[J]. Precision Agriculture, 2016, 17: 365-374.
- [10] 罗汝英. 江西省低山区杉林土壤障碍性条件的判别分析[J]. 土壤学报, 1980, 17(2): 145-155.
Luo R Y. Discriminant analysis of inhibiting soil conditions of Chinese fir stands in low mountain district of Jiangxi Province[J]. Acta Pedologica Sinica, 1980, 17(2): 145-155.
- [11] 熊国炎. 两组线性判别分析在土壤分类中的应用[J]. 土壤, 1980,

- (5): 177–181.
Xiong G L. Two groups of linear discriminant analysis applied for the soil classification[J]. *Soils*, 1980(5): 177–181.
- [12] Brus D J, Bogaert P, Heuvelink G B. Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information[J]. *European Journal of Soil Science*, 2008, 59(2): 166–177.
- [13] 邱琳, 李安波, 赵玉国. 基于Fisher判别分析的数字土壤制图研究[J]. *土壤通报*, 2012, 43(6): 1281–1286.
Qiu L, Li A B, Zhao Y G. Digital soil mapping based on Fisher discriminant analysis[J]. *Chinese Journal of Soil Science*, 2012, 43(6): 1281–1286.
- [14] 吴海波, 赵晓慎, 王治国, 张超. 基于Bayes判别分析模型的水土保持区划[J]. *水土保持科学*, 2012, 10(2): 88–91.
Wu H B, Zhao X Z, Wan Z G, Zhang C. Regional division for soil and water conservation based on Bayes discriminant analysis model[J]. *Science of Soil and Water Conservation*, 2012, 10(2): 88–91.
- [15] 鲁如坤. 土壤农业化学分析方法[M]. 北京: 中国农业科技出版社, 2000: 146–196.
Lu R K. Soil agricultural chemical analysis method[M]. Beijing: China Agricultural Science and Technology Press, 2000: 146–196.
- [16] Andrew R W, Keith D C (王萍译). 统计模式识别(第三版)[M]. 北京: 电子工业出版社, 2015: 163–201.
Andrew R W, Keith D C (Translated by Wang P). Statistical pattern recognition (3rd ed.) [M]. Beijing: Electronic Industry Press, 2015: 163–201.
- [17] 李娟, 章明清, 孔庆波, 等. 构建县域早稻氮磷钾施肥的系统聚类方法研究[J]. *植物营养与肥料学报*, 2017, 23(2): 531–538.
Li J, Zhang M Q, Kong Q B, *et al.* Building fertilization categories of N, P and K fertilization for early rice using systematic clustering method in county territory[J]. *Journal of Plant Nutrition and Fertilizer*, 2017, 23(2): 531–538.
- [18] 袁志发, 宋世德. 多元统计分析(第二版)[M]. 北京: 科学出版社, 2009: 278–293.
Yuan Z F, Song S D. Multivariate statistical analysis (2nd ed.)[M]. Beijing: Science Press, 2009: 278–293.
- [19] 李柏年, 吴礼斌. MATLAB数据分析方法[M]. 北京: 机械工业出版社, 2015: 81–104.
Li P N, Wu L B. MATLAB data analysis method [M]. Beijing: Mechanical Industry Press, 2015: 81–104.
- [20] 毛达如, 张承东. 多点肥料效应函数的动态聚类方法[J]. *北京农业大学学报*, 1991, 17(2): 49–54.
Mao D R, Zhang C D. Dynamic clustering method of multipoint fertilizer response function[J]. *Acta Agriculturae Universitatis Pekinensis*, 1991, 17(2): 49–54.
- [21] 章明清, 林仁坝, 林代炎, 姜永. 极值判别分析在三元肥效模型推荐施肥中的作用[J]. *福建农业学报*, 1995, 10(2): 54–59.
Zhang M Q, Lin R X, Lin D Y, Jiang Y. Function of distinguish analysis on extreme value in recommendatory fertilization for three-fertilizer efficiency model[J]. *Fujian Journal of Agricultural Sciences*, 1995, 10(2): 54–59.
- [22] 周鸣铮. 测土施肥的科学基础[J]. *土壤通报*, 1984, 15(4): 156–160.
Zhou M Z. Scientific foundation of soil testing and recommendation fertilization[J]. *Chinese Journal of Soil Science*, 1984, 15(4): 156–160.
- [23] 周鸣铮. 中国的测土施肥[J]. *土壤通报*, 1987, 18(1): 7–13.
Zhou M Z. Soil testing and recommendation fertilization in China[J]. *Chinese Journal of Soil Science*, 1987, 18(1): 7–13.
- [24] 金耀青. 配方施肥的方法及其功能—对我国配方施肥工作的述评[J]. *土壤通报*, 1989, 20(1): 46–49.
Jin Y Q. Formula fertilization methods and its functions—review of formula fertilization work in China[J]. *Chinese Journal of Soil Science*, 1989, 20(1): 46–49.
- [25] 陆允甫, 吕晓男. 中国测土施肥工作的进展与展望[J]. *土壤学报*, 1995, 32(3): 241–251.
Lu Y P, Lü X N. Progress and prospect in fertilizer recommendation based on soil testing in China[J]. *Acta Pedologica Sinica*, 1995, 32(3): 241–251.